Sparse Bayesian Learning-based Channel Estimation for IRS-aided Millimeter Wave Massive MIMO Systems

Agrim Agarwal, Amrita Mishra and Priyanka Das
International Institute of Information Technology Bangalore (IIITB), India 560100
{agrim.agarwal, amrita.mishra, priyanka.das}@iiitb.ac.in

***Abstract*—Intelligent reflecting surface (IRS)-aided millimeter wave (mmWave) systems are potential contenders for large scale deployment in 5G and beyond communication. Owing to the passive nature of IRS systems, the acquisition of accurate channel state information in bandwidth constrained scenarios becomes a highly challenging task. This work leverages the inherent sparsity associated with cascaded channels of IRS mmWave systems to develop novel sparse Bayesian learning (SBL)-based channel estimation algorithms. The underlying row-wise temporal correlation of the effective angular domain sparse channel matrix is first theoretically demonstrated followed by the development of the temporal SBL channel estimation approach. Further, utilizing the scaling property of cascaded channels for multiple users, low complex variants of the proposed SBL solutions are developed. Simulation results demonstrate significant improvement of the proposed SBL schemes over existing techniques.**

***Index Terms*—Intelligent reflecting surface (IRS), millimeter wave (mmWave), sparse Bayesian learning (SBL), temporal correlation, scaling property of cascaded channels**

I. INTRODUCTION

Intelligent reflecting surface (IRS)-aided millimeter wave (mmWave) systems are well-suited for 5G and beyond communication owing to their high spectral and energy efficiency [1]. IRS comprises of a large number of passive reflecting elements which introduce appropriate phase shifts in incident signals, resulting in their constructive addition at the receiver. Thus, IRS enables the enhancement of communication range and received power without any additional energy requirement. However, the high gains of IRS-based communication systems can be realized using accurate channel estimates.

To this end, the passive nature of the reflective elements in IRS makes channel estimation a highly challenging task that can be carried out only at the base station (BS) or user equipment (UEs). A conventional method involves turning on-off IRS elements in a sequential manner to estimate a single BS-IRS-user channel at a time [2]. This method suffers from a large pilot overhead with increase in the number of IRS elements. The work in [3] utilizes the quasi-static nature of the BS-IRS channel to develop a two-timescale approach wherein the BS-IRS channel is estimated over a long period of time and the IRS-user channels over much shorter intervals. Although this approach significantly reduces the pilot overhead compared to [2], it suffers from increased computational complexity and poor estimation accuracy due to the full-duplex mode of operation at the BS. Further, it can be observed that none of the above approaches consider the inherent sparsity in mmWave channels which can be suitably leveraged for accurate channel estimation in low pilot overhead scenarios.

In this regard, the work in [4] considers a virtual angular domain-based sparse representation of the IRS mmWave channel. The authors therein employ a greedy compressed sensing technique i.e. orthogonal matching pursuit (OMP) for channel estimation, resulting in poor performance for a lower number of pilots. In [5], the authors have utilized the common BS-IRS channel between different users to develop efficient sparsity-based channel estimation techniques. However, for severely ill-posed estimation scenarios i.e. the number of reflecting elements in IRS significantly greater than the number of pilots, the estimation accuracy of such approaches is observed to deteriorate drastically. Next, a sparse IRS channel estimation approach based on a hybrid combination of expectation propagation approximation (EPA) and generalized approximate message passing (GAMP) is proposed in [6]. However, the authors therein consider spatial sparsity in the channel only at the BS side while ignoring it at the IRS, thereby making the channel model too simplistic for practical consideration.

Motivated by the shortcomings of existing works, the contributions of this paper can be summarized as follows:

- A novel sparse Bayesian learning (SBL) approach is proposed to estimate IRS-based cascaded channels in severely bandwidth constrained scenarios. The scheme is based on a parameterized prior assignment and employs the expectation and maximization (EM) algorithm for hyperparameter-based cascaded channel estimation.
- The temporal correlation associated with active rows of the effective angular domain sparse channel matrix is first theoretically demonstrated followed by development of the temporal SBL approach.
- Efficient and low-complex counterparts of the proposed SBL algorithms are developed based on the scaling property of cascaded channels for multiple users.

The paper is organized into five sections. Section II outlines the IRS-based mmWave massive multiple-input multiple-output (MIMO) system model along with the sparse angular

978-1-6654-6483-3/23/$31.00 ©2023 IEEE

978-1-6654-6483-3/23/$31.00 ©2023 IEEE

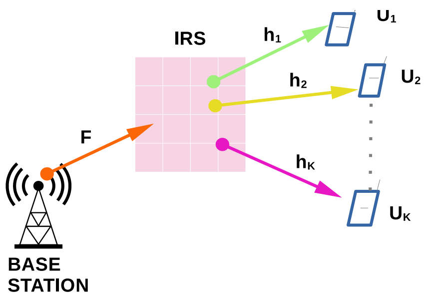

Fig. 1: IRS-aided multi-user massive MIMO system.

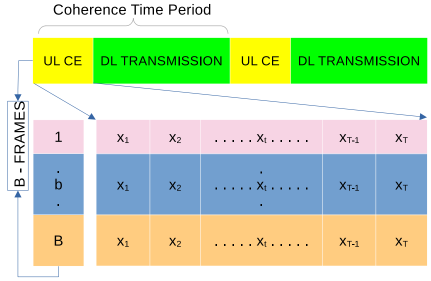

Fig. 2: B copies of T Length Pilots

domain channel matrix. Section III discusses the proposed SBL-based schemes and their computationally efficient counterparts. Simulation results and concluding remarks are presented in sections IV and V, respectively.

The notations followed are summarized as follows. Scalar values are shown using small case letters x. Vectors and matrices are denoted by small case boldface letters $(\mathbf{x})$ and uppercase boldface letters $(\mathbf{X})$, respectively. The notation $\mathbf{x}_{l\cdot}$ represents the l-th row of the matrix $\mathbf{X}$. The estimate of the variable $\mathbf{x}$ at the jth iteration is denoted by the superscript notation $\mathbf{x}^{(j)}$ and $\hat{\mathrm{x}}$ denotes the estimate of the x. The notation $\mathbf{I}_t$ represents an $t \times t$ identity matrix. A complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is represented by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For a matrix $\mathbf{X}$, vec$(\mathbf{X})$ represents vector obtained by stacking the columns of the matrix $\mathbf{X}$. The symbol $\otimes$ represents the standard matrix Kronecker product. $\|\mathbf{X}\|_2^2$ and $\|\mathbf{X}\|_{\mathcal{F}}$ represent the l_2 norm and Frobenius norm of the matrix $\mathbf{X}$ respectively.

II. System Model

A MIMO multi-user wireless system operating in a time division duplex (TDD) mode is considered where M base station (BS) antennas serve K single antenna users using L reflecting elements of an IRS. Each IRS element is assumed to be a point source that aggregates and reflects the BS signal to the users. A diagrammatic representation of the multi-user MIMO IRS system is demonstrated in Fig. 1. Let the reflection coefficient vector $\mathbf{c} \in \mathbb{C}^{L\times 1}$ associated with the IRS be given by $\mathbf{c} = [c_1, c_2, \ldots, c_L]^T$ where $c_l = e^{j\rho_l}$ and ρ_l denote the complex reflection coefficient and phase shift corresponding to the l-th IRS element, respectively. Let the MIMO channel matrices from the BS to the IRS and IRS to the k-th user be denoted by $\mathbf{F} \in \mathbb{C}^{L\times M}$ and $\mathbf{h}_k^H \in \mathbb{C}^{1\times L}$, respectively. The cascaded channel matrix $\mathbf{G}_k \in \mathbb{C}^{L\times M}$ is defined as

$$\mathbf{G}_k = diag(\mathbf{h}_k^H)\mathbf{F}. \tag{1}$$

Let the channel coherence time period be divided into two parts i.e. an uplink channel estimation phase followed by the data transmission on the downlink. The BS requires perfect channel state information (CSI) of the downlink cascaded channel for data transmission. Hence, to conserve channel capacity as well as to reduce processing load at the users, uplink channel estimation is performed by leveraging the channel reciprocity property of the BS-IRS-User link, similar to related works such as [1], [2], [5]. For the uplink channel estimation, the cascaded channel $\mathbf{G}_k$ is estimated at the BS for all K users. The k-th user transmits a T length known sequence $\mathbf{s}_k = [s_{k,1}, s_{k,2}, \ldots, s_{k,T}]^H \in \mathbb{C}^{T\times 1}$ which constitutes a pilot frame. The pilot sequences $\mathbf{s}_k \ \forall \ 1 \leqslant k \leqslant K$ are assumed to be mutually orthogonal such that the interference between transmissions corresponding to different users is nullified [5]. Further, the k-th user transmits B such copies of the pilot frame $\mathbf{s}_k$. The received symbol matrix at the BS $\mathbf{Y}_b \in \mathbb{C}^{M\times T}$ corresponding to the b-th frame can be expressed as

$$\mathbf{Y}_b = \sum_{k=1}^{K} \mathbf{F}^H \left(diag\left(\mathbf{c}_b\right)\right) \mathbf{h}_k \mathbf{s}_k^H + \mathbf{N}_b \tag{2}$$

$$= \sum_{k=1}^{K} \mathbf{F}^H \left(diag\left(\mathbf{h}_k\right)\right) \mathbf{c}_b \mathbf{s}_k^H + \mathbf{N}_b \tag{3}$$

$$= \sum_{k=1}^{K} \mathbf{G}_k^H \mathbf{c}_b \mathbf{s}_k^H + \mathbf{N}_b, \tag{4}$$

where the reflection coefficient vector for the b-th frame $\mathbf{c}_b = [c_{b,1}, c_{b,2}, \ldots, c_{b,L}]^H \in \mathbb{C}^{L\times 1}$ and $\mathbf{N}_b \in \mathbb{C}^{M\times T}$ represents the complex circularly symmetric additive white Gaussian noise matrix such that $\mathbf{N}_b \sim \mathcal{CN}\left(\mathbf{0}_{M\times T}, \sigma^2 \mathbf{I}_M\right)$. Consider $\mathbf{y}_{b,k} \triangleq \frac{1}{PT}\mathbf{Y}_b \mathbf{s}_k \in \mathbb{C}^{M\times 1}$, where P denotes the user transmit power. It can be simplified as

$$\mathbf{y}_{b,k} = \frac{1}{PT}\left[\sum_{k=1}^{K} \mathbf{G}_k^H \mathbf{c}_b \mathbf{s}_k^H + \mathbf{N}_b\right]\mathbf{s}_k \tag{5}$$

$$= \mathbf{G}_k^H \mathbf{c}_b + \mathbf{n}_{b,k}, \tag{6}$$

with $\mathbf{n}_{b,k} = \frac{1}{PT}\mathbf{N}_b \mathbf{s}_k \in \mathbb{C}^{M\times 1}$. The simplification from (5) to (6) can be attributed to the orthogonal nature of $\mathbf{s}_k$. Concatenating the received vectors $\mathbf{y}_{b,k}$ of the k-th user for all B frames, one obtains $\mathbf{Y}_k = [\mathbf{y}_{1,k}, \mathbf{y}_{2,k}, \ldots, \mathbf{y}_{B,k}] \in \mathbb{C}^{M\times B}$. The quantities $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_B] \in \mathbb{C}^{L\times B}$ and $\mathbf{N}_k = [\mathbf{n}_{1,k}, \mathbf{n}_{2,k}, \ldots, \mathbf{n}_{B,k}] \in \mathbb{C}^{M\times B}$ denote similar concatenated matrices. The overall system equation for the k-th user can be written as

$$\mathbf{Y}_k = \mathbf{G}_k^H \mathbf{C} + \mathbf{N}_k. \tag{7}$$

Further, the cascaded channel $\mathbf{G}_k$ follows a scaling property such that [5]

$$\mathbf{G}_k = \mathbf{A}_k \mathbf{G}_1, \tag{8}$$

where $\mathbf{G}_1$ denotes the cascaded channel of the first user and the scaling coefficient matrix $\mathbf{A}_k \in \mathbb{C}^{L\times L}$ is given by

$$\mathbf{A}_k = diag\left(\left[\frac{\left[\boldsymbol{h}_k^H\right]_1}{\left[\boldsymbol{h}_1^H\right]_1}, \frac{\left[\boldsymbol{h}_k^H\right]_2}{\left[\boldsymbol{h}_1^H\right]_2}, \ldots, \frac{\left[\boldsymbol{h}_k^H\right]_L}{\left[\boldsymbol{h}_1^H\right]_L}, \right]\right). \quad (9)$$

Without loss of generality, one can rewrite (8) and (9) for any arbitrary reference user. Using (8), one can rewrite (7) as

$$\mathbf{Y}_k^H = \mathbf{C}^H \mathbf{A}_k \mathbf{G}_1 + \mathbf{N}_k^H. \quad (10)$$

Let I and J denote the total number of active spatial paths between BS-IRS and IRS-user respectively. The BS and IRS employ uniform linear arrays (ULA) for signal transmission and reception [7]. Considering the physical propagation model similar to mmWave-related works such as [5], [8], [9], the channel matrices $\mathbf{h}_k$ and $\mathbf{F}$ can be expressed as

$$\mathbf{F} = \sqrt{\frac{LM}{I}} \sum_{i=1}^{I} \alpha_i \mathbf{a}_L\left(\theta_i^{AoA}\right) \mathbf{a}_M^H\left(\phi_i^{AoD}\right), \quad (11)$$

$$\mathbf{h}_k = \sqrt{\frac{L}{J}} \sum_{j=1}^{J} \beta_{k,j} \mathbf{a}_L\left(\delta_{k,j}^{AoD}\right), \quad (12)$$

where $\alpha_i \sim \mathcal{CN}\left(0, \sigma_{\alpha_i}^2\right)$ and $\beta_{k,j} \sim \mathcal{CN}\left(0, \sigma_{\beta_{k,j}}^2\right)$ are complex channel gains associated with the i-th spatial path between BS-IRS and j-th spatial path between IRS-k-th user respectively. The quantities θ_i^{AoA}, ϕ_i^{AoD} represent the angle of arrival at IRS and angle of departure from BS respectively for i-th active spatial path between BS-IRS, whereas $\delta_{k,j}^{AoD}$ represents the angle of departure for j-th active spatial path from IRS to k-th user. The notation $\mathbf{a}_Q(\Theta)$ used above represents an array steering vector of length Q and spatial angle (AoA/AoD) Θ given by $\mathbf{a}_Q(\Theta) = \frac{1}{\sqrt{Q}}\left[1, e^{j\frac{2\bar{d}}{\lambda}\pi\sin\Theta}, \ldots, e^{j\frac{2\bar{d}}{\lambda}\pi(Q-1)\sin\Theta}\right]^H \in \mathbb{C}^{Q\times 1}$ where d and λ denote the inter-antenna element spacing and carrier wavelength, respectively. Without loss of generality, the inter-antenna element spacing d is set as $d = \lambda/2$. On substituting (11) and (12) in (1), the cascaded channel matrix $\mathbf{G}_k$ can now be written as

$$\mathbf{G}_k = \sum_{i=1}^{I}\sum_{j=1}^{J} \sqrt{\frac{L^2 M}{IJ}} \alpha_i \beta_{k,j} \mathbf{a}_L\left(\delta_{k,j}^{AoD} + \theta_i^{AoA}\right) \mathbf{a}_M^H\left(\phi_i^{AoD}\right). \quad (13)$$

Note that $\left(\delta_{k,j}^{AoD} + \theta_i^{AoA}\right)$ denotes an effective spatial angle between the IRS and k-th user and will be referred to as the effective AoA at IRS/user throughout the manuscript. Typically, the exact number of spatial paths and AoAs/AoDs associated with $\mathbf{G}_k$ are unknown at the BS. For estimation of the cascaded channel matrix, an alternate virtual angle domain representation can be employed similar to related works such as [9]. Consider G_R and G_T number of angular grid points which provide suitably segregated effective AoA and AoD spaces, respectively. The spatial angles $\mathcal{D}_R = \{D_r : D_r \in [-\pi/2, \pi/2]\ ,\ 1 \leqslant r \leqslant G_R\}$ in effective AoA space and $\mathcal{D}_T = \{D_t : D_t \in [-\pi/2, \pi/2], 1 \leqslant t \leqslant G_T\}$ in AoD space at base station(BS) are defined such that $\sin(D_r) = \left(\frac{2}{G_R}(r-1) - 1\right)\ \forall\ 1 \leqslant r \leqslant G_R$ and $\sin(D_t) = \left(\frac{2}{G_T}(t-1) - 1\right)\ \forall\ 1 \leqslant t \leqslant G_T$ respectively. The dictionary matrices $\mathbf{U} \in \mathbb{C}^{L\times G_R}$ and $\mathbf{V} \in \mathbb{C}^{M\times G_T}$ are defined as

$$\mathbf{U} = \left[\mathbf{a}_L(D_1), \ldots, \mathbf{a}_L(D_{G_R})\right], \quad (14)$$

$$\mathbf{V} = \left[\mathbf{a}_M(D_1), \ldots, \mathbf{a}_M(D_{G_T})\right], \quad (15)$$

where the array steering vectors of spatial angles $D_r \in \mathcal{D}_R$ and $D_t \in \mathcal{D}_T$ represent the columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. One can observe that the spatial angles in $\mathcal{D}_R$ and $\mathcal{D}_T$ may not necessarily contain the actual AoDs and effective AoAs associated with $\mathbf{G}_k$, respectively, resulting in a mismatch. However, simulation results interestingly demonstrate that there is no discernible deterioration in the performance of the proposed scheme resulting from this mismatch. Thus, using (14) and (15), one can approximate $\mathbf{G}_k$ in the virtual angular domain as

$$\mathbf{G}_k \approx \mathbf{U}\mathbf{X}_k\mathbf{V}^H. \quad (16)$$

Note that the total number of active spatial paths constituting the cascaded channel $\mathbf{G}_k$ in (13) is significantly smaller than G_R and G_T. As a result of this, $\mathbf{X}_k \in \mathbb{C}^{G_R\times G_T}$ represents an angular domain sparse matrix with very few non-zero entries. The (r,t)-th non-zero entry denotes the complex cascaded channel gain for the active spatial paths from the r-th effective AoA at IRS/user and t-th AoD at BS. Further, considering hermitian on both sides of (7) the system model for the k-th user can be expressed as $\mathbf{Y}_k^H = \mathbf{C}^H\mathbf{G}_k + \mathbf{N}_k^H$, where it is important to note that conventional estimation approaches like least squares (LS) will require $\mathbf{C}^H$ to be a full rank matrix, i.e $B \geqslant L$. In scenarios with a large number of reflective elements in the IRS, one needs to increase the size of pilot overhead resulting in spectral inefficiency. Recent works on sparse channel estimation in mmWave systems [8], [9] demonstrate that spatial sparsity can be successfully leveraged to obtain accurate channel estimates for ill-posed estimation scenarios i.e. $B \ll L$, in the IRS-aided mmWave massive MIMO context. Therefore, using (16), one can rewrite the system model as

$$\mathbf{Y}_k^H = \mathbf{C}^H\mathbf{U}\mathbf{X}_k\mathbf{V}^H + \mathbf{N}_k^H. \quad (17)$$

Let $\bar{\mathbf{X}}_k = \mathbf{X}_k\mathbf{V}^H \in \mathbb{C}^{G_R\times M}$ denote the effective angular domain cascaded channel matrix. The equation (17) can be simplified as

$$\mathbf{Y}_k^H = \mathbf{C}^H\mathbf{U}\bar{\mathbf{X}}_k + \mathbf{N}_k^H. \quad (18)$$

It can be verified that $\bar{\mathbf{X}}_k = \mathbf{X}_k\mathbf{V}^H \in \mathbb{C}^{G_R\times M}$ is a row-sparse matrix i.e. it comprises of few non-zero rows which correspond to the active spatial paths from the effective AoAs of IRS/user. Thus, estimation of the unknown cascaded channel matrices $\mathbf{G}_k$ for all K users now reduces to the estimation of their row-sparse counterparts $\bar{\mathbf{X}}_k$. Using the scaling property in (8) and (10), (17) can be rewritten as

$$\mathbf{Y}_k^H = \mathbf{C}^H\mathbf{A}_k\mathbf{U}\mathbf{X}\mathbf{V}^H + \mathbf{N}_k^H, \quad (19)$$

where the cascaded channel matrix for an arbitrary reference user can be denoted as $\mathbf{U}\mathbf{X}\mathbf{V}^H = \mathbf{G}$. For the sake of simplicity, it is set as $\mathbf{G} = \mathbf{G}_1 = \mathbf{U}\mathbf{X}_1\mathbf{V}^H$ similar to (8). Similarly, $\bar{\mathbf{X}}_1 = \mathbf{X}_1\mathbf{V}^H \in \mathbb{C}^{G_R \times M}$. Employing this, (19) can be rewritten as

$$\mathbf{Y}_k^H = \mathbf{C}^H\mathbf{A}_k\mathbf{U}\bar{\mathbf{X}}_1 + \mathbf{N}_k^H. \tag{20}$$

It can be observed from (20) that estimation of $\mathbf{G}_k$ for the k-th user can alternately be visualized as estimation of its unknown scaling factor matrix $\mathbf{A}_k$ and $\bar{\mathbf{X}}_1$ i.e. the effective angular domain cascaded channel matrix for the first user (also reference user here). Thus, using (20), one needs to estimate the scaling coefficients $\mathbf{A}_k$ for all the K users and $\bar{\mathbf{X}}_1$ thereby significantly reducing the estimation load.

Optimization of IRS reflection coefficients requires knowledge of the cascaded channel matrices of all users [10]. Channel estimation further becomes challenging owing to the passive nature of IRS [3], [6]. With an increase in IRS elements, conventional schemes suffer from a high pilot overhead. This work leverages the inherent sparsity associated with massive MIMO multi-user IRS systems to develop accurate channel estimation schemes with low pilot overhead.

III. Proposed Sparse Bayesian Learning-Based Algorithms

This section develops two categories of channel estimation schemes based on the system models in (18) and (20).

A. Multiple Sparse Bayesian Learning (MSBL)-based Cascaded Channel Estimation

The proposed SBL-based scheme is applied for estimation of the row sparse matrix $\bar{\mathbf{X}}_k$ in (18) $\forall\ 1 \leqslant k \leqslant K$ users. The subscript k is dropped in the following discussion for notational brevity. Unlike conventional sparse recovery techniques such as FOCUSS [11] which assigns a fixed sparsity promoting prior, the multiple sparse Bayesian learning framework considers a parameterized Gaussian prior of the form $p\left(\bar{\boldsymbol{x}}_{l.};\gamma\left(l\right)\right) \triangleq \mathcal{CN}\left(\mathbf{0}_{M\times 1},\gamma\left(l\right)\mathbf{I}_M\right)$ corresponding to each row $\bar{\boldsymbol{x}}_{l.}$, $1 \leqslant l \leqslant G_R$, of $\bar{\mathbf{X}}$. Assuming the rows to be independent, the overall prior of $\bar{\mathbf{X}}$ can be expressed as

$$p\left(\bar{\mathbf{X}};\boldsymbol{\gamma}\right) = \prod_{l=1}^{G_R} p\left(\bar{\boldsymbol{x}}_{l.};\gamma\left(l\right)\right), \tag{21}$$

where $\boldsymbol{\gamma} = \left[\gamma\left(1\right),\gamma\left(2\right),\ldots,\gamma\left(G_R\right)\right]^T \in \mathbb{C}^{G_R\times 1}$ denotes an unknown hyperparameter vector such that as $\gamma\left(l\right) \to 0$, the l^{th} row of $\bar{\mathbf{X}}$ i.e. $\bar{\mathbf{x}}_{l.} \to 0$. The posterior density of $\bar{\mathbf{X}}$ can be computed as [9], [12] $p\left(\bar{\mathbf{X}} \mid \mathbf{Y};\boldsymbol{\gamma}\right) \sim \mathcal{CN}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$, where,

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^H\mathbf{Y}^H \in \mathbb{C}^{G_R\times M}, \tag{22}$$

$$\boldsymbol{\Sigma} = \left(\sigma^{-2}\boldsymbol{\Psi}^H\boldsymbol{\Psi} + \left(\boldsymbol{\Gamma}\right)^{-1}\right)^{-1} \in \mathbb{C}^{G_R\times G_R}, \tag{23}$$

denote the *a posteriori* mean and covariance matrices respectively with $\boldsymbol{\Psi} = \mathbf{C}^H\mathbf{U} \in \mathbb{C}^{B\times G_R}$ and $\boldsymbol{\Gamma} = diag\left(\boldsymbol{\gamma}\right) \in \mathbb{C}^{G_R\times G_R}$. As evident from (23), the maximum *a posteriori* (MAP) estimate of $\bar{\mathbf{X}}$ is dependent on the unknown hyperparameter matrix $\boldsymbol{\Gamma}$. Thus, the estimation of $\hat{\bar{\mathbf{X}}}$ is analogous to the estimation of the unknown hyperparameter vector $\boldsymbol{\gamma}$. A conventional maximum likelihood (ML) estimate of $\boldsymbol{\gamma}$ can be obtained as $\hat{\gamma}_{ML} = \arg\max_{\boldsymbol{\gamma}} \log p\left(\mathbf{Y};\boldsymbol{\gamma}\right)$. However, solving the above maximization problem is mathematically intractable [11]. Therefore, the expectation maximization (EM) algorithm is adopted for $\hat{\gamma}$ estimation similar to related works such as [12]. The expectation (E-step) in the t-th iteration involves evaluation of average log-likelihood $\mathcal{L}\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(t-1)}\right)$ given by

$$\mathcal{L}\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(t-1)}\right) = \mathrm{E}_{\bar{\mathbf{X}}\mid\mathbf{Y};\boldsymbol{\gamma}^{(t-1)}}\left\{\log p\left(\mathbf{Y},\bar{\mathbf{X}};\boldsymbol{\gamma}\right)\right\}. \tag{24}$$

Further, the maximization (M-step) maximizes (24) with respect to $\boldsymbol{\gamma}$, and can be expressed as

$$\begin{aligned}\boldsymbol{\gamma}^{(t)} \equiv \arg\max_{\boldsymbol{\gamma}} \mathrm{E}_{\bar{\mathbf{X}}\mid\mathbf{Y};\boldsymbol{\gamma}^{(t-1)}}\left\{\log p\left(\mathbf{Y} \mid \bar{\mathbf{X}};\boldsymbol{\gamma}\right)\right\} \\ + \mathrm{E}_{\bar{\mathbf{X}}\mid\mathbf{Y};\boldsymbol{\gamma}^{(t-1)}}\left\{\log p\left(\bar{\mathbf{X}};\boldsymbol{\gamma}\right)\right\}.\end{aligned} \tag{25}$$

Ignoring the term $\log p\left(\mathbf{Y} \mid \bar{\mathbf{X}};\boldsymbol{\gamma}\right)$ in (25) since it is independent of $\boldsymbol{\gamma}$, one can rewrite (25) as

$$\boldsymbol{\gamma}^{(t)} \equiv \arg\max_{\boldsymbol{\gamma}} \mathrm{E}_{\bar{\mathbf{X}}\mid\mathbf{Y};\boldsymbol{\gamma}^{(t-1)}}\left\{\log p\left(\bar{\mathbf{X}};\boldsymbol{\gamma}\right)\right\}. \tag{26}$$

Finally, the hyperparameter update for the t-th iteration is given as [9]

$$\gamma^{(t)}\left(l\right) = \boldsymbol{\Sigma}_{l,l}^{(t)} + \frac{1}{M}\left|\boldsymbol{\mu}_{l.}^{(t)}\right|^2 \forall 1 \leqslant l \leqslant G_R, \tag{27}$$

where $\boldsymbol{\Sigma}_{l,l}^{(t)}$ represents the (l,l) element of the covariance matrix $\boldsymbol{\Sigma}^{(t)}$ and $\boldsymbol{\mu}_{l.}^{(t)}$ represents the l-th row of the *a posteriori* mean in the t-th iteration. The *a posteriori* matrices $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ are evaluated using (22) and (23), respectively by considering $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{(t-1)} = \mathrm{diag}(\boldsymbol{\gamma}^{(t-1)})$. The proposed MSBL algorithm is executed for a maximum number of K_{max}^{MSBL} EM iterations to attain desirable convergence i.e $\|\boldsymbol{\Gamma}^{(t)} - \boldsymbol{\Gamma}^{(t-1)}\|_2 < \epsilon_M$, where, t = K_{max}^{MSBL} and ϵ_M is a small value. The estimate of $\bar{\mathbf{X}}$ after convergence is given by $\hat{\bar{\mathbf{X}}} = \boldsymbol{\mu}^{K_{max}^{MSBL}}$. Thus, the

Algorithm 1 Proposed scaling factor-based TMSBL channel estimation technique

Input: $\mathbf{Y}_k, \mathbf{U}, \boldsymbol{\Pi}, \mathbf{C}, T_{max}, \mathbf{A}_k^{(0)}$
Initialization: $\boldsymbol{\Pi}^{(0)} = \mathbf{I}_{G_R\times G_R}$, $\boldsymbol{\Phi} = \left[\mathbf{C}^H\mathbf{A}_1^{(0)}\mathbf{U};\mathbf{C}^H\mathbf{A}_2^{(0)}\mathbf{U};\ldots;\mathbf{C}^H\mathbf{A}_K^{(0)}\mathbf{U}\right] \in \mathbb{C}^{BK\times G_R}$, $t = 1$
while $t < T_{max}$ or $Tr\|\boldsymbol{\Pi}^{(t)} - \boldsymbol{\Pi}^{(t-1)}\| < 10^{-2}$ **do**
  E-step: Evaluate *a posteriori* mean $\boldsymbol{\chi}^{(t)}$ and covariance $\boldsymbol{\Xi}^{(t)}$ using (35) and (36) for $user - 1$
  M-step: Obtain hyperparameter and covariance matrix update from (37), (38) and (39)
  $\hat{\mathbf{X}}_1^{(t)} = \boldsymbol{\chi}^{(t)}$
  Update $\mathbf{A}_k^{(t)}$ for $2 \leqslant k \leqslant K$ using $\boldsymbol{\chi}^{(t)}$ and $\mathbf{Y}_k$ as shown in (40)
  $t \leftarrow t + 1$
end while
return: $\hat{\mathbf{G}}_k = \hat{\mathbf{A}}_k U\hat{\mathbf{X}}_1$ where $\hat{\mathbf{A}}_k = \mathbf{A}_k^{(t)}$ and $\hat{\mathbf{X}}_1 = \boldsymbol{\chi}^{(t)}$

MSBL-based cascaded channel estimate can be evaluated as $\hat{\mathbf{G}}_{MSBL} = \mathbf{U}\hat{\bar{\mathbf{X}}}$.

B. Temporal Correlation-based MSBL (TMSBL) for Cascaded Channel Estimation

One can leverage the row-wise temporal correlation in $\bar{\mathbf{X}}$ to further improve the MSBL-based cascaded channel estimation. Details about the temporal correlation can be found in the Appendix section. A positive definite matrix $\mathbf{B}^l \in \mathbb{C}^{M\times M}$ containing the correlation information of the l-th row of $\bar{\mathbf{X}}$ is considered. To avoid overfitting, $\mathbf{B}^1 = \cdots = \mathbf{B}^l = \ldots \mathbf{B}^{G_R} = \mathbf{B}$ is assumed [13]. The parameterized prior of $\bar{\mathbf{X}}$ in terms of the hyperparameter vector $\boldsymbol{\gamma}$ and covariance matrix $\mathbf{B}$, can now be expressed as

$$p\left(\bar{\mathbf{X}};\boldsymbol{\gamma},\mathbf{B}\right) = \prod_{l=1}^{G_R} p\left(\bar{\mathbf{x}}_{l.};\gamma\left(l\right),\mathbf{B}\right), \tag{28}$$

where, $p\left(\bar{\boldsymbol{x}}_{l.};\gamma\left(l\right),\mathbf{B}\right) \triangleq \mathcal{CN}\left(0,\gamma\left(l\right)\mathbf{B}\right)$. The E-step based on (24) and (28) can be expressed as

$$\begin{aligned} &\mathcal{L}\left(\boldsymbol{\gamma}^{(t)},\mathbf{B}^{(t)} \mid \boldsymbol{\gamma}^{(t-1)},\mathbf{B}^{(t-1)}\right) && (29)\\ =\ & \mathrm{E}_{\bar{\mathbf{X}}|\mathbf{Y};\boldsymbol{\gamma}^{(t-1)},\mathbf{B}^{(t-1)}}\left\{\log p\left(\mathbf{Y},\bar{X};\boldsymbol{\gamma},\mathbf{B}\right)\right\} && (30)\\ =\ & \mathrm{E}_{\bar{\mathbf{X}}|\mathbf{Y};\boldsymbol{\gamma}^{(t-1)},\mathbf{B}^{(t-1)}}\left\{\log p\left(\mathbf{Y}\mid\bar{\mathbf{X}};\boldsymbol{\gamma},\mathbf{B}\right)\right\} + \\ & \mathrm{E}_{\bar{\mathbf{X}}|\mathbf{Y};\boldsymbol{\gamma}^{(t-1)},\mathbf{B}^{(t-1)}}\left\{\log p\left(\bar{\mathbf{X}};\boldsymbol{\gamma},\mathbf{B}\right)\right\}. && (31)\end{aligned}$$

This is followed by the M-Step for $\boldsymbol{\gamma}$ and $\mathbf{B}$. Ignoring the first term in (31) independent of $\boldsymbol{\gamma}$ and $\mathbf{B}$, the M-step estimation for $\boldsymbol{\gamma}$ can be written as

$$\boldsymbol{\gamma}^{(t)} \equiv \arg\max_{\boldsymbol{\gamma}} \mathrm{E}_{\bar{\mathbf{X}}|\mathbf{Y};\boldsymbol{\gamma}^{(t-1)},\mathbf{B}^{(t-1)}}\left\{\log p\left(\bar{\mathbf{X}};\boldsymbol{\gamma},\mathbf{B}\right)\right\}. \tag{32}$$

The hyperparameter update in the t-th iteration can be obtained by taking gradient of (32) with respect to $\boldsymbol{\gamma}$ and applying approximations used by the authors in [13] as

$$\gamma^{(t)}\left(l\right) = \boldsymbol{\Sigma}_{l,l}^{(t)} + \frac{1}{M}\boldsymbol{\mu}_{l.}^{(t)}\left(\mathbf{B}^{-1}\right)^{(t-1)}\boldsymbol{\mu}_{l.}^{H(t)}, \tag{33}$$

$\forall\ 1 \leqslant l \leqslant G_R$. The *a posteriori* mean and covariance matrices $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ respectively, are evaluated using (22) and (23), respectively, by considering $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{(t-1)} = \mathrm{diag}(\boldsymbol{\gamma}^{(t-1)})$ as evaluated in (33) in the previous $(t-1)$-th iteration. Similarly, one can obtain an update of the covariance matrix $\mathbf{B}$ in the t-th iteration as [13]

$$\mathbf{B}^{(t)} = \frac{\tilde{\boldsymbol{B}}^{(t)}}{\|\tilde{\boldsymbol{B}}^{(t)}\|_{\mathcal{F}}}, \tag{34}$$

where $\tilde{\mathbf{B}}^{(t)} = \sum_{l=1}^{G_R}\frac{\boldsymbol{\mu}_{l.}^{(t)}\cdot\boldsymbol{\mu}_{l.}^{H(t)}}{\gamma^{(t)}(l)} + \eta\mathbf{I}$ and η denotes a regularization parameter added to make the algorithm more robust in low to medium SNR ranges. The hyperparameter and covariance matrix updates in (33) and (34) are repeated until convergence criteria of the hyperparameter estimate $\hat{\boldsymbol{\Gamma}}$ is achieved. Similar to the proposed MSBL approach, the TMSBL algorithm is executed for a maximum number of K_{max}^{TMSBL} EM iterations to attain desirable convergence i.e $\|\boldsymbol{\Gamma}^{(t)} - \boldsymbol{\Gamma}^{(t-1)}\|_2 < \epsilon_T$, where, t = K_{max}^{TMSBL} and ϵ_T represents a small constant.

C. Scaling factor-based TMSBL (STMSBL)

The MSBL and TMSBL schemes can be modified based on the scaling factor-based formulation in (20) to develop their computationally efficient cascaded channel estimation counterparts. The scaling factor MSBL (SMSBL) approach can be derived similar to the scaling factor TMSBL (STMSBL) algorithm developed in this section. Considering (20), it is important to observe that the cascaded channel estimation for all K users now corresponds to estimation of the scaling factor matrices $\mathbf{A}_k,\ \forall\ 1 \leqslant k \leqslant K$ and the row sparse matrix $\bar{\mathbf{X}}_1$ of the reference user. Let $\boldsymbol{\Phi} = \left[\mathbf{C}^H\mathbf{A}_1\mathbf{U};\mathbf{C}^H\mathbf{A}_2\mathbf{U};\ldots;\mathbf{C}^H\mathbf{A}_K\mathbf{U}\right] \in \mathbb{C}^{BK\times G_R}$ and $\tilde{\mathbf{Y}} = \left[\mathbf{Y}_1^H;\mathbf{Y}_2^H;\ldots;\mathbf{Y}_K^H\right] \in \mathbb{C}^{BK\times M}$. Let the hyperparameter vector estimate of the reference user in the $(t-1)$-th iteration be denoted as $\boldsymbol{\varrho}^{(t-1)} \in \mathbb{C}^{G_R\times 1}$ and $\boldsymbol{\Pi}^{(t-1)} = diag\left(\varrho^{(t-1)}\left(1\right),\ldots,\varrho^{(t-1)}\left(G_R\right)\right) \in \mathbb{C}^{G_R\times G_R}$. The posterior density of the unknown row sparse matrix is only calculated for the reference user. The *a posteriori* mean and covariance matrices in the t-th iteration corresponding to $\bar{\mathbf{X}}_1$ can be evaluated as

$$\boldsymbol{\chi}^{(t)} = \sigma^{-2}\boldsymbol{\Xi}^{(t-1)}\boldsymbol{\Phi}^H\tilde{\mathbf{Y}} \in \mathbb{C}^{G_R\times M}, \tag{35}$$

$$\boldsymbol{\Xi}^{(t)} = \left(\sigma^{-2}\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \left(\boldsymbol{\Pi}^{(t-1)}\right)^{-1}\right)^{-1} \in \mathbb{C}^{G_R\times G_R} \tag{36}$$

Similar to the steps followed in (32) - (34), the EM updates for $\boldsymbol{\Pi}$ and covariance matrix $\bar{\mathbf{B}}$ of the reference user can be obtained by substituting the mean and covariance values evaluated in (35) and (36) as,

$$\varrho^{(t)}\left(l\right) = \boldsymbol{\Xi}_{l,l}^{(t)} + \frac{1}{M}\boldsymbol{\chi}_{l.}^{(t)}\left(\bar{\mathbf{B}}^{(t-1)}\right)^{-1}\boldsymbol{\chi}_{l.}^{(t)H}, \tag{37}$$

$$\tilde{\bar{\boldsymbol{B}}}^{(t)} = \sum_{l=1}^{G_R}\frac{\boldsymbol{\chi}_{l.}^{(t)H}\cdot\boldsymbol{\chi}_{l.}^{(t)}}{\varrho^{(t)}\left(l\right)} + \eta\mathbf{I}, \tag{38}$$

$$\bar{\boldsymbol{B}}^{(t)} = \frac{\tilde{\bar{\boldsymbol{B}}}^{(t)}}{\|\tilde{\bar{\boldsymbol{B}}}^{(t)}\|_{\mathcal{F}}}. \tag{39}$$

The scaling coefficient matrix $\mathbf{A}_k$ can be readily evaluated in the t-th iteration using the least squares approach as [5]

$$\hat{\mathbf{A}}_{\mathbf{k}}^{(t)} = diag\left(\left(\mathbf{H}^{:,\Omega_A}\right)^{\dagger}\left[\bar{\mathbf{Q}}_k\right]\right), \tag{40}$$

where $\bar{\mathbf{Q}}_k = vec\left(\mathbf{Y}_K^H\right) \in \mathbb{C}^{BM\times 1}$, $\mathbf{H} = \left(\mathbf{U}\boldsymbol{\chi}^{(t)}\right)^T \otimes \mathbf{C}^H$, and $\Omega_A = \{i + (i-1)\,L \mid i \in \{1,2,\ldots,L\}\}$ represents the set of indices corresponding to the L non zero diagonal entries in $\mathbf{A}_k$ scaling coefficient matrix. The notation $(\cdot)^{\dagger}$ represents Moore-Penrose pseudoinverse matrix. The EM updates of the algorithm are repeated till convergence $\|\boldsymbol{\Pi}^{(t)} - \boldsymbol{\Pi}^{(t-1)}\| < \epsilon_{ST}$ is achieved or for a maximum number of EM iterations $t = K_{max}^{STMSBL}$, whichever happens earlier. At the end of EM iterations the estimated value of $\bar{\mathbf{X}}_1$ is given as $\hat{\bar{\mathbf{X}}}_1 = \boldsymbol{\chi}^t$. The final cascaded channel estimate for k-th user can be evaluated as $\hat{\mathbf{G}}_k = \hat{\mathbf{A}}_k\mathbf{U}\hat{\bar{\mathbf{X}}}_1$. A good initial estimate of the scaling coefficients $A_k(l,l),\ \forall\ 1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K$ can be obtained as follows. The cascaded channels $\mathbf{G}_k$ corresponding to all $1 \leqslant k \leqslant K$ users are initially estimated using the proposed

MSBL algorithm. Employing the relationship in (8), one can obtain the initial estimate as $\mathbf{A}_k^{(0)}(l,l) = \frac{1}{M}\sum_{m=1}^{M}\frac{\hat{G}_k^{MSBL}(l,m)}{\hat{G}_1^{MSBL}(l,m)}$. The computational complexity of the proposed STMSBL scheme can be obtained as $\mathcal{O}\left(K_{max}^{STMSBL}KL^2MB\right)$. A detailed analysis and computational complexity comparison with existing cascaded channel estimation techniques is deferred to a future extension of this work.

IV. Simulation Results

A mmWave massive MIMO system is considered with $M = 64$ BS antennas, $L = 64$ IRS reflecting elements for $K = 4$ single antenna users, where $T = K$ length pilot sequences are transmitted for $B = 30$ frames. A total number of 100 Monte Carlo iterations are considered for the numerical experiment. The IRS reflection coefficients are set as $|C_{l,b}| = 1, \forall\ 1 \leqslant l \leqslant L$ and $\forall\ 1 \leqslant b \leqslant B$, which are assumed to be known at the BS. A total number of $I = 8$ and $J = 1$ spatial paths is considered between the BS-IRS and IRS-user, respectively. For the purposes of simulations, $\sigma_{\alpha_i}^2 = 1$ and $\sigma_{\beta_{k,j}}^2 = 1\ \forall i, \forall j, \forall k$. The AoAs and AoDs θ_i^{AoA}, ϕ_i^{AoD} and $\delta_{k,j}^{AoD}$ are randomly generated from $\mathcal{U}\left(-\pi/2, \pi/2\right)$. A total number of $G_R = G_T = 128$ spatial grid points are assumed. The stopping criteria of the proposed schemes MSBL, TMSBL, SMSBL, and STMSBL are fixed as $K_{max}^{MSBL} = 200$, $K_{max}^{TMSBL} = 400$, $K_{max}^{SMSBL} = 800$ and $K_{max}^{STMSBL} = 1000$, respectively with $\epsilon_M = 10^{-5}$, $\epsilon_T = 10^{-4}$, $\epsilon_{SM} = 5\times10^{-3}$ and $\epsilon_{ST} = 10^{-2}$. The performance of the proposed cascaded channel estimation schemes are compared with existing works in terms of normalized mean squared error (NMSE) evaluated as NMSE $= \frac{\|\hat{\boldsymbol{G}}_k - \mathbf{G}_k\|_2^2}{\|\mathbf{G}_k\|_2^2}$. Simulations results are demonstrated to compare the proposed techniques with existing works such as subspace multi-user joint channel estimation (SMJCE) [5] , simultaneous OMP (SOMP) [14] and subspace-based SOMP (S-SOMP) [8] and GENIE - aided [5] method. The GENIE - aided method assumes complete knowledge of all the AoAs and AoDs for channel estimation and thus serves as a performance benchmark.

Fig 3(a) illustrates the NMSE comparison of the proposed and existing approaches for an ill-posed estimation scenario which considers $L = 64$ IRS elements and $B \ll L$ i.e $B = 30$ frames. It can be observed that the proposed SMSBL and STMSBL schemes outperform the existing SMJCE and other SOMP-based methods. The channel estimation in STMSBL and SMSBL approaches reduces to estimation of the reference user's row sparse matrix $\bar{\mathbf{X}}_1$ and scaling factors of remaining $K-1$ users. Thus, the number of unknowns to be estimated is significantly less in comparison to MSBL and TMSBL resulting in a performance improvement. The NMSE of the SMJCE approach increases at high SNR due to an SNR dependent penalty factor in the objective function which yields erroneous results for ill-posed estimation scenarios.

Fig 3(b) illustrates the NMSE performance comparison of the proposed and existing schemes for a range of B values and fixed $L = 64$, at an SNR of $= 15$dB. It can be observed that the proposed SBL-based schemes demonstrate high level of accuracy for a lower number of frames corresponding to bandwidth constrained scenarios. For $B = 20$, the SMSBL and STMSBL approaches yield an NMSE of 10^{-3} and 5×10^{-4} respectively, whereas the existing SMJCE scheme yields a higher NMSE of 10^{-1}. Further, the performance of the existing SMJCE method becomes comparable to that of the SMSBL and STMSBL approaches for $B \geqslant 40$.

Fig 3(c) presents a performance comparison of different cascaded channel estimation approaches for a range of L values with $B = 30$ and SNR$=$ 15dB. It can be observed that for $L = 30$, i.e for $L = B$ the NMSE associated with the SMJCE, SMSBL, and STMSBL schemes are comparable with each other. However, for $L = 90$. i.e when $L \gg B$, the NMSE of the SMSBL and STMSBL approaches are significantly lower than SMJCE technique. Therefore, one can conclude that the degradation of performance with increase in L is gradual for the proposed SBL approaches while a steep increase is observed in case of the existing SMJCE scheme.

V. Conclusion

Novel SBL-based schemes have been developed for channel estimation in IRS-aided massive MIMO mmWave systems by leveraging the inherent spatial sparsity of mmWave channels. Improvements in the proposed SBL approaches have been designed by exploiting the row-wise temporal correlation of the effective angular domain matrix and a multi-user scaling property of cascaded IRS channels. Simulation results demonstrate the superiority of the proposed schemes over existing approaches in terms of channel estimation accuracy for low pilot overhead estimation scenarios.

VI. Acknowledgment

This work was supported by funds from IIITB COMET Foundation, established under the Advanced Communication Systems vertical of the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS), Department of Science and Technology, Govt. of India.

Appendix

The row sparse matrix $\bar{\mathbf{X}} = \mathbf{X}\mathbf{V}^H$ where $\mathbf{V}$ denotes a known dictionary matrix of the form (15) and $\mathbf{X}$ represents a sparse angular domain matrix with non-zero entries corresponding to cascaded channel gains associated with active effective AoAs/AoDs. One can express $\bar{\mathbf{X}} = [\bar{x}_{r,m}]\ \ \forall 1 \leqslant r \leqslant G_R, 1 \leqslant m \leqslant M$, $\mathbf{V} = [v_{m,t}]\ \ \forall 1 \leqslant m \leqslant M,\ 1 \leqslant t \leqslant G_T$ and $\mathbf{X} = [x_{r,t}]\ \ \forall\ 1 \leqslant r \leqslant G_R, 1 \leqslant t \leqslant G_T$. Consider the l-th row of $\bar{\mathbf{X}}$ i.e. $\bar{\mathbf{x}}_{l.}$ to be non zero. To establish row-wise temporal correlation in $\bar{\mathbf{X}}$, one needs to prove that for any two elements of the l-th row say $\bar{x}_{l,p}$ and $\bar{x}_{l,q}$, $Cov\{\bar{x}_{l,p}\bar{x}_{l,q}\} = E\{\bar{x}_{l,p}\bar{x}_{l,q}^*\} - E\{\bar{x}_{l,p}\}E\{\bar{x}_{l,q}^*\} \neq 0$, where $E\{.\}$ denotes the standard expectation operator. Using $\bar{\mathbf{X}} = \mathbf{X}\mathbf{V}^H$, one can rewrite $\bar{x}_{l,p} = \sum_{j=1}^{G_T} x_{l,j}v_{p,j}^*$. Taking expectation of $\bar{x}_{l,p}$ yields

$$E\{\bar{x}_{l,p}\} = \sum_{j=1}^{G_T} E\{x_{l,j}\}v_{p,j}^*. \tag{41}$$

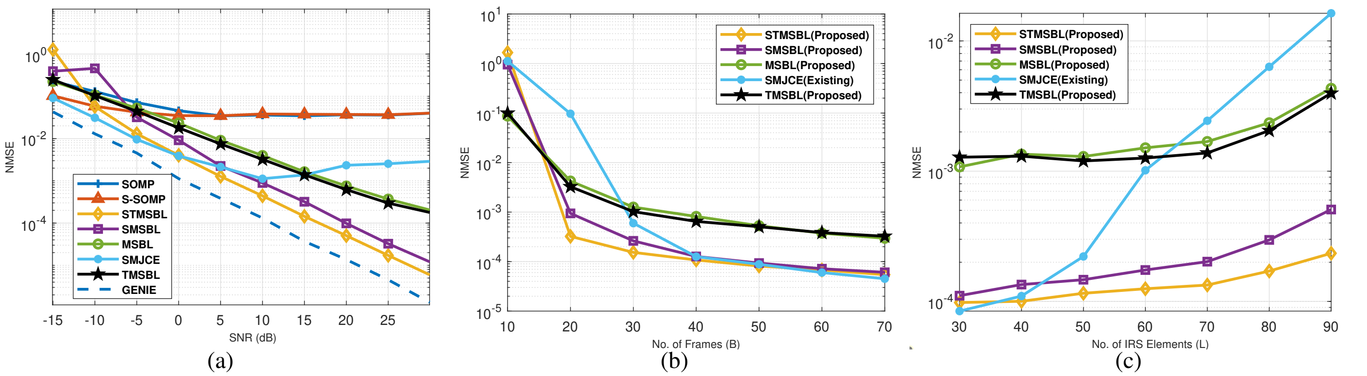

Fig. 3: For $M = 64$ BS antennas and $K = 4$ users (a) NMSE performance comparison of channel estimate versus SNR with $B = 30, L = 64$ (b) NMSE performance comparison versus no. of frames (B) with $L = 64$, SNR $= 15$dB. (c) NMSE performance comparison versus no. of reflecting elements in IRS (L) with $B = 30$ and SNR $= 15$dB.

Similarly, $E\{\bar{x}_{l,q}\} = \sum_{j=1}^{G_T} E\{x_{l,j}\}v_{q,j}^*$. Consider a set $\boldsymbol{\vartheta}$, with cardinality $|\boldsymbol{\vartheta}| = Q$ which contains indices corresponding to the non-zero entries of $\mathbf{x}_{l\cdot}$ such that $Q \ll G_T$. As mentioned earlier in (16), the non-zero entries of the angular domain sparse matrix $\mathbf{X}$ correspond to the complex cascaded channel gains associated with the active spatial paths. Without loss of generality, let the complex cascaded channel gains corresponding to the non-zero entries of $\mathbf{x}_{l\cdot}$ be distributed as $\zeta_z \sim \mathcal{CN}\left(0, \sigma_{\zeta_z}^2\right)\ \forall\ 1 \leqslant z \leqslant Q$. Thus, $E\{\bar{x}_{l,p}\} = \sum_{u\in\boldsymbol{\vartheta}} E\{\zeta_u\}v_{p,u}^* = 0$ and $E\{\bar{x}_{l,q}\} = \sum_{w\in\boldsymbol{\vartheta}} E\{\zeta_w\}v_{q,w}^* = 0$ since $E\{\zeta_u\} = E\{\zeta_w\} = 0\ \forall\ u, w \in \boldsymbol{\vartheta}$. Hence, $Cov\{\bar{x}_{l,p}\bar{x}_{l,q}\} = E\{\bar{x}_{l,p}\bar{x}_{l,q}^*\}$. Substituting the expressions for $\bar{x}_{l,p}$ and $\bar{x}_{l,q}$ in $E\{\bar{x}_{l,p}\bar{x}_{l,q}^*\}$, one can write

$$Cov\{\bar{x}_{l,p}\bar{x}_{l,q}\} = E\left\{\left[\sum_{u\in\boldsymbol{\vartheta}} \zeta_u v_{p,u}^*\right]\left[\sum_{w\in\boldsymbol{\vartheta}} \zeta_w^* v_{q,w}\right]\right\} \tag{42}$$

$$= E\left\{\sum_{(u=w)} |\zeta_u|^2 v_{p,u}^* v_{q,u}\right\} + E\left\{\sum_{\substack{u\in\boldsymbol{\vartheta} \\ (u\neq w)}}\sum_{w\in\boldsymbol{\vartheta}} \zeta_u\zeta_w^* v_{p,u}^* v_{q,w}\right\}. \tag{43}$$

The second term in (43) employs the independence associated with cascaded gains corresponding to different spatial paths and simplifies as $E\left\{\sum_{\substack{u\in\boldsymbol{\vartheta} \\ (u\neq w)}}\sum_{w\in\boldsymbol{\vartheta}} \zeta_u\zeta_w^* v_{p,u}^* v_{q,w}\right\} = 0$. The first term of (43) can be simplified as $E\left\{\sum_{(u=w)} |\zeta_u|^2 v_{p,u}^* v_{q,u}\right\} = \sum_{(u=w)\in\boldsymbol{\vartheta}} \sigma_{\zeta_u}^2 v_{p,u}^* v_{q,w}$. Thus, $Cov\{\bar{x}_{l,p}\bar{x}_{l,q}\} = \sum_{(u=w)\in\boldsymbol{\vartheta}} \sigma_{\zeta_u}^2 v_{p,u}^* v_{q,w}$. Therefore, using the above result, it can be easily demonstrated that the covariance matrix $\mathbf{B}^l = E\{(\bar{\mathbf{x}}_{l\cdot})^H(\bar{\mathbf{x}}_{l\cdot})\} = [Cov\{\bar{x}_{l,p}\bar{x}_{l,q}\}]$ $\forall\ 1 \leqslant p \leqslant M, 1 \leqslant q \leqslant M$ associated with the l-th row of $\bar{\mathbf{X}}$.

References

[1] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—part I: Fundamentals, solutions, and future opportunities," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1398–1402, 2021.

[2] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *ICASSP - IEEE Internat. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, 2020, pp. 5000–5004.

[3] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. on Commun.*, vol. 69, no. 11, pp. 7736–7747, 2021.

[4] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.

[5] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmwave mimo systems," *IEEE Trans. on Wirel. Commun.*, pp. 1–1, 2023.

[6] L. Zhou, J. Dai, W. Xu, and C. Chang, "Sparse channel estimation for intelligent reflecting surface assisted massive mimo systems," *IEEE Trans on Green Commun. and Net.*, vol. 6, no. 1, pp. 208–220, 2022.

[7] H. Han, Y. Liu, and L. Zhang, "On half-power beamwidth of intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1333–1337, 2021.

[8] C.-R. Tsai, Y.-H. Liu, and A.-Y. Wu, "Efficient compressive channel estimation for millimeter-wave large-scale antenna systems," *IEEE Trans. on Signal Process.*, vol. 66, no. 9, pp. 2414–2428, 2018.

[9] A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Sparse Bayesian learning-based channel estimation in millimeter wave hybrid MIMO systems," in *2017 IEEE 18th Int. Workshop on Signal Process. Adv. in Wireless Commun. (SPAWC)*, 2017, pp. 1–5.

[10] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. on Signal Process.*, vol. 70, pp. 3964–3977, 2022.

[11] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. on Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[12] R. Prasad and C. R. Murthy, "Bayesian learning for joint sparse ofdm channel estimation and data detection," in *2010 IEEE Global Telecommun. Conf. GLOBECOM 2010*, 2010, pp. 1–6.

[13] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE Journal of Selected Topics in Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[14] J. Tropp, A. Gilbert, and M. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proceedings. (ICASSP '05). IEEE Int. Conf. on Acoustics, Speech, and Signal Process., 2005.*, vol. 5, 2005, pp. 721–724.